LINDA L. HILL

Spatial Access to, and Display of, Global Change Data: Avenues for Libraries

Geospatial retrieval systems are considered in the framework of the U.S. Global Change Data and Information System (GCDIS), a cooperative effort of a set of federal agencies that have vast amounts of data and information relevant to the global change user community. The characteristics of an ideal geospatial retrieval system are proposed and five existing systems are described as examples of the types of spatial retrieval available today. Existing systems fall far short of the ideal system envisioned but show exciting promise. Librarians can participate in the evaluation of these systems and in the evolving development of GCDIS through the Library Information Subgroup of the Global Change Data Management Working Group.

INTRODUCTION

Spatial access to, and display of, global change data is addressed here in the context of the United States Global Change Research Program (USGCRP) and, in particular, of the Global Change Data and Information System (GCDIS). The USGCRP was established to observe, understand, and predict global change and to make its results available for use in policy matters. The goals of the data management activities of the USGCRP are to archive, preserve, and make data available in a useful form now and in the future for researchers and any other groups who can benefit from access to these data. The challenge of the data management program is to handle the massive amounts of highly diverse data and information generated and used by global change research activities. These data, and accompanying information products, reflect multidisciplinary global change issues that include human and natural forcing factors, ecological change, biodiversity, human interactions, and comprehensive assessments.

The federal agencies involved in the USGCRP have cooperated to organize the Global Change Data and Information System. The GCDIS builds on existing and planned agency-mission-responsive data and information systems that have information relevant to the global change user community and links these systems to each other and to users. The enhanced interoperability of the agency systems is achieved through standards, common approaches, technology sharing, and data policy

coordination. The design focuses on providing service for a diverse community of global change users—researchers, policymakers, educators, private industry, and the public. The objective is to aid the user community in learning what data and information are available, in having the key holdings available in useful forms with ready access, and in being assured of their quality and continued availability. Multiple levels of access and service are included in GCDIS, from phone and mail and fax to high bandwidth transfer of image data. The strategy for this initiative is described in the *U.S. Global Change Data and Information System Implementation Plan* (Committee on the Environment and Natural Resources, 1994). Other key documents from this effort are included in the references at the end of this discussion.

The Global Change Data Management Working Group (GCDMWG)[1] is building the GCDIS. Agencies participating in GCDIS are identifying their resources and developing coordinated ways for a wider audience than the primary research community to find and use their data and information. The GCDIS is being developed and implemented by the participating agencies with a minimum of additional infrastructure. The first interface effort was the beta version of the GCDIS gopher, registered in the Spring of 1994. In the Spring of 1995, the gopher is being revised on the basis of the "lessons learned" from the use of the beta version, and a web home page is being developed. These developments are being done through these agencies: the gopher is mounted at the National Oceanic and Atmospheric Administration (NOAA) by the Environmental Services Division Information Management staff, and NASA staff has contributed to its design; the DOE's Carbon Dioxide Information Analysis Center (CDIAC) staff is developing the home page; the developments are being coordinated through the Executive Secretariat of the GCDMWG that consists of only two people: the Executive Secretary Les Meredith and myself. The working group itself is structured as follows:

- A Contacts Group guides the overall coordination effort for data management and three subgroups. It consists of representatives from the participating agencies. A Principals Group, which is the chief agency representative to the Working Group, is convened for budgetary decisions. There is no direct funding for the GCDIS effort. All activities are supported through contributions from the agencies. The Contacts Group is chaired by Dixon Butler of NASA's Mission to Planet Earth Program.
- The Content Subgroup's responsibility is to coordinate efforts to identify and set priorities for GCDIS data and information and related information products to resolve user needs, identify information gaps, encourage the participation of researchers in GCDIS, serve as a point

of contact for users that need special products, and coordinate the establishment of data and information quality assessment processes. The Content Subgroup is chaired by Michael Farrell of the Department of Energy.

- The Access Subgroup "strives to create a seamless search, browse, order, and delivery of global change data and information from many agencies" (from draft GCDIS implementation plan, March 1995), promotes the use of the Z39.50 protocol and emerging Internet tools, and is taking an active role in bringing together the players in other related federal initiatives: the Federal Geographic Data Committee (FGDC), the National Environmental Data Inventory (NEDI), and the Government Information Locator Service (GILS). The development of the GCDIS gopher and the home page is being done through the Access Subgroup. The Global Change Research Information Office (GCRIO), which provides help desk support to GCDIS, is linked to this subgroup. The Access Subgroup is chaired by Tom Mace of the Environmental Protection Agency.
- The third subgroup is the Library Information Subgroup (LIS). GCDIS may be the only high level federal government interagency data activity that specifically brings library and information center resources and expertise to the table at this level. The responsibility of the Library Information Subgroup is to support the Working Group efforts by:

  A. building an infrastructure of libraries and librarians for GCDIS implementation;
  B. evaluating GCDIS from a library user's perspective and providing user needs analysis;
  C. linking data resources to information resources for knowledge management;
  D. promoting GCDIS to libraries and developing approaches to user education;
  E. advising on data and information processing standards and systems from the library perspective (from draft LIS implementation plan, March 1995). [2]

  The Library Information Subgroup is chaired by Roberta Rand of the National Agricultural Library.

Last year each subgroup began sponsored pilot projects: the Contents Subgroup is investigating methodologies for determining priorities of research interest and thus content acquisition using the trace gas research community as the testbed. The Access Subgroup is sponsoring the Assisted Search for Knowledge (ASK) project that is developing a client-server search, retrieval, and use system based on a semantic network

search engine. This will be discussed later here because the ASK project has a geographic information system component. The Library Information Subgroup is sponsoring a GCDIS evaluation project in the state of Virginia that involves public school (elementary, upper elementary, and high school), community college, university, public library, and community groups in the use of GCDIS as it exists and involves them in the further design and enhancement of the system. This is known as the Library Access, Search, and Retrieval (LASR) Pilot Project and is being led by the Computer Sciences Department of the University of Virginia.

The data and information resources that are within the purview of the GCDIS effort include both those resulting from focused Global Change Research Project (GCRP)-funded research and the data and information resulting from other agency efforts. The potential resources to be tapped include (from the draft implementation plans of the participating agencies submitted in the Fall of 1994):

- terabytes of data derived from Environmental Protection Agency (EPA) regulatory activities;
- more than two dozen data and information service nodes in the Department of Commerce, including the satellite and in situ data of the National Oceanic and Atmospheric Administration; the holdings of the NOAA Central Library; the statistical, economic, and demographic information of the Bureau of the Census; and the technical, engineering, and business information available through the National Technical Information Service (NTIS);
- data and information from the Department of Energy's Atmospheric Radiation Measurement (ARM) Archive, the Carbon Dioxide Information Analysis Center, the Oak Ridge National Laboratory (ORNL) Distributed Active Archive Center (DAAC), their Office of Scientific and Technical Information (OSTI), the Energy Information Administration (EIA), and the Energy Efficiency and Renewable Energy Network (EREN);
- data archives from the Department of Agriculture's Agricultural Research Service (ARS); Cooperative State Research, Education and Extension Service (CSREES); their Economic Research Service (ERS); the Forest Service (FS); and the Terrestrial Ecosystems Regional Research and Analysis (TERRA) Laboratory as well as the holdings of the National Agricultural Library (NAL) and the World Agricultural Outlook Board (WAOB);
- the terabyte data archives of the National Science Foundation's National Center for Atmospheric Research (NCAR), including the NCAR library holdings, and the Long-Term Ecological Research Network;
- the data servers of three Department of Interior programs: The Global Change Research Program Data Server, the United States GS Earth

Resources Observation Systems (EROS) Data Center, and the National Biological Service (NBS) Global Change Research Program Data Management Center as well as additional data centers and data systems that archive many types of geospatial data (e.g., water quality measurements, forest inventory plots, soil sampling, wetland maps, digital cartographic data files, census counts, and timber inventories);

- the Global Change Master Directory (GCMD) and the Earth Observation System Data and Information System (EOSDIS) from the National Aeronautics and Space Administration (NASA) as well as data and information from the Pathfinder program, the Crustal Dynamics Data Information System (CDDIS), and NASA's Scientific and Technical Information Program.

Most of the agencies, in addition to the resources listed here, have their own data-sharing agreements or links to state, regional, and worldwide resources. There are also other nodes such as NASA's Consortium for International Earth Science Information Network (CIESIN) which provides access to socioeconomic data and information as well as scientific data and information and the Global Change Research Information Office which provides user services to those accessing this wealth of information.

Each participating agency has agreed to a set of minimum requirements and has prepared its own GCDIS implementation plan. Each subgroup has also prepared an implementation plan targeting activities for the next three years.

This brief introduction is intended to give the reader a picture of the vast amount of national and international data and information resources within the purview of the GCDIS and the interagency structure established to provide coordinated access to these resources. This whole effort is not happening in isolation, by any means. The National Academy of Sciences reviews and approves the Working Group plans and through overlapping committee memberships, reciprocal presentations and discussions, and focused coordination meetings, the GCDIS effort is harmonized with similar programs. The GCDIS effort seeks to be an entry point to distributed global change data and information for diverse audiences and particularly to make the data and information accessible by and through libraries.

There are many substantive issues in relation to GCDIS: priorities for content acquisition; data management priorities; interoperability, accessibility; management of a distributed system; political and budgetary issues for a multiagency effort; resource discovery structures and mechanisms for a distributed system; metadata standards; meeting the needs of a broad range of users; and the international, multiformat, and interdisciplinary aspects of the system. The question that will be explored

here is focused on the extent to which GCDIS, in its present status, can provide access to the available data and information by geographic location through spatial access, display, and manipulation. Although the emphasis in GCDIS is on global data sets and information and many of the problems in this area require global data gathering (e.g., global warming, ozone depletion, El Niño, etc.), GCDIS data and information can often be linked to specific geographic areas through latitude and longitude coordinates and geographic place names.

Before illustrating the current status of spatial access to federal government global change data and information, some of the characteristics of an ideal spatial retrieval system for finding georeferenced information will be outlined.

SPATIAL ACCESS—WHAT WOULD WE REALLY LIKE TO BE ABLE TO DO?

As background, my primary interest is in information retrieval. The research that was done for my dissertation (Hill, 1990) compared the effectiveness of the retrieval of georeferenced information using word-based descriptions (place names) and spatial representations. The research was based on ninety-nine geoscience articles about the Mediterranean region, each having a sketch map of the research area. Spatial representations of the maps contained in the articles were used as the "true" representations of the geographic study areas. This was done with the GRASS GIS software[3] and resulted in raster representations of the sketch maps in the articles. For every document in the set, the other ninety-eight documents were ranked by geographic similarity on the basis of their spatial overlap or distance away. Rankings between documents were also determined based on geographic terminology, both assigned descriptors and the text of the bibliographic citations and abstracts. Comparison between spatially-based rankings and word-based rankings were used to determine how successfully the words in the bibliographic records reflect the geographic relationships among the documents. The results showed that there was only a weak correlation between the rankings based on geographic words and the rankings based on spatial map-based representations.

The study revealed other confirmations of the ineffectiveness of words for the retrieval of georeferenced information. Using spatial overlap as a sign of relevance between one document and another, word-based retrieval was only able to average 50 percent recall and 41 percent precision—that is, only 50 percent of the relevant documents were identified (on average) and only 41 percent of the documents retrieved by geographic words were overlapping areas on average. These recall and precision figures are typical of other studies of the effectiveness of text-based retrieval; this

impreciseness of our retrieval tools is not limited to geographic terminology. But in the case of the geographic "aboutness" of documents, we do have an alternative—a spatial representation of the areas.

Some of the bibliographic records used in the study were from GeoRef[4] and contained latitude and longitude coordinates to describe the study areas. There were obvious errors in a few of these due to such errors as reversing top and bottom latitudes or conversion errors when the records were loaded on the Orbit system. When seen graphically, the errors are obvious, but these errors are very difficult to detect when looking at a string of latitude and longitude values. The other problem with the system of assigning two latitudes and two longitudes to describe an area—still the dominant approach—is the impreciseness of this regular polygon parallel to the equator when describing an irregular area. However, the rankings based on latitude and longitude representations did have a higher correlation to the spatial representations than the words in the study.

My ideal system for finding and retrieving georeferenced information will have text documents, data sets, maps, photographs, and all other forms of information that are geocoded to represent the geographic coverage areas of the items, formatted in accordance with widely accepted interoperable metadata standards. The users of this system would be able to retrieve all forms of information that relate to particular geographic areas from distributed databases.

The geocoding describes geographic areas—not just a latitude and longitude point but the representation of an area of coverage or study. The areal representation should closely match the study area of the document or data set and not be grossly generalized. For example, a data set or technical report covering the Gulf Coast of Florida should not be represented by a box that also takes in most of the Gulf of Mexico. A river valley should not be represented by a point at the mouth of the river nor by a box that greatly exceeds the environs of the river valley. The representation of the area should be close enough to the actual geographic area for appropriate differentiation for information retrieval. This does not require the exactness of a map representation but, in an ideal system, geographic representation would not be limited to either single points representing the geographic center of an area or to regular polygons that are parallel to the equator.

The spatial query capability in the ideal system would be imbedded in a retrieval system that supports other types of query parameters: free-text words and phrases from the documents and metadata; subject headings and descriptors; dates and date ranges; authors; organizational affiliations and sponsorship; format; type of information; and other technical details. The spatial query capability would present the user with a map of

the world. Through pointing and zooming, the user zeros in on the area of interest and draws a shape that surrounds the area of interest. The user is not limited to a rectangular shape but can instead outline the Gulf Coast of Florida or the Mississippi River valley; an oil producing region or a fault zone; a forest tract or a park; a biological zone or an ocean current. If simple geographic names or census tract numbers or political entities or other easily named areas will suffice for the query, then the user can simply enter these names with the assurance that the system can, if requested to, translate the place names into spatial definitions and expand the search to those items in the database that may not have those specific names attached to them but that include the area nevertheless.

To bridge the gap between data that are geocoded with latitude and longitude coordinates (or a scheme that can be translated to latitude/longitude descriptions) and documents that have their geographic "aboutness" described with place names, there needs to be a translation scheme. This would be a database that defines place names spatially with latitude and longitude coordinates so that the information-retrieval system can prompt users for place names that are in a designated geographic search area or expand a search that includes place names to their spatial equivalents. In this way, users of the information-retrieval system can use either coordinates or place names and have the option of expanding their search to the other form of description. There is some preliminary work being done in this area and a great deal more needs to be done.

The matching mechanisms for this ideal retrieval system will compare the user's query area with the spatial coverage areas of the items in the database (or databases) and return a ranked list based on a metric, such as a similarity measure.

- For overlapping areas (i.e., the query area and the item area overlap), the area of the overlap times 2 would be divided by the sum of the query area plus the area of the item that was matched. This process adjusts for widely differing sizes between the query and the target item in the database as well as the amount of overlap. The highest similarity value comes from the largest overlap and the most similar sizes. A global data set would have a low similarity to the Florida Gulf Coast even though it does contain it with this scheme. All database items with areas overlapping the query area would be ranked by the system and returned in descending order by the strength of the similarity.
- Database items whose geographic areas share a common boundary with the query area but do not overlap it would have a similarity value of zero.
- Database items that do not overlap the query area and do not have a common boundary would be related to the query area by a negative number representing the distance between the closest boundaries.

The user would be allowed to set a similarity cut off value so that only overlapping areas are retrieved (i.e., similarity greater than zero) or only database items with a "significant" amount of overlap.

Probably the ideal system will provide various ways of limiting or filtering retrieved sets of data and information by spatial extent. A metric could be used that indicates the percentage of the query area covered by the database items so that complete coverage could be distinguished from partial coverage. Doug Nebert of the USGS and Project Alexandria at the University of California, Santa Barbara, is also considering other ways of giving the user a sense of the spatial extent of the database items in relation to the query area. One idea is to compute the "square degrees" of an area by multiplying the latitude and longitude dimensions of spatial descriptions. With this artificial measure, areas could be divided into rough sets such as "local," "regional," "continental," and "global" and the user given a choice of the extent of coverage she or he wants to have in the retrieved set.

The system would return a ranked list of the database items that meet all of the search criteria with the geographic component of the search expressed in spatial terms. Exactly how this is to be done has not received much attention, as far as I know. The user would now like to ask the system to display the items by their geographic areas on the map used for the query and show the types of items by a visual key: data sets, images, full text, metadata only, books, articles, technical reports, maps, photographs, etc. From this display, items can be chosen for review and selection—for downloading or ordering.

Now with the data in hand, the user will need further system help to manipulate the data sets and images selected. This will involve browsing snapshots of the images (or progressive displays of the images that the user can stop when she or he has seen enough[5]) before downloading them. Full documentation for the data should be available so that it is clear how the data are registered and encoded; the circumstances under which the data were obtained; the instruments used; the quality control applied; etc. Eventually, the user may want to display and manipulate the data sets with a Geographic Information System (GIS) and correlate data from different sources.

In summary, the ideal system will present a graphic spatial query mode for geographically related searches; it will return a ranked list of relevant items based on a similarity measure or other metric; it will display the resulting "hits" on a map with proper identification so that items can be reviewed for selection; and it will provide tools for browsing images and manipulating data sets.

FIVE EXAMPLE SYSTEMS FROM FEDERAL AGENCIES WITH SPATIAL RETRIEVAL

The Global Change Data and Information System provides an initial starting point to a rich, diverse, and distributed set of scientific and social data and information resources. The major contributors to GCDIS were introduced earlier in this discussion. Participation is not limited to these

seven agencies, however. Any federal agency with data and information that can contribute to an understanding of global change is being encouraged to participate in GCDIS, and the agencies' international, state, local, regional resources will also be incorporated. For the purpose of representing the spatial access, display, and manipulation capabilities of this diverse group, the following systems will be highlighted:

- The GC-ASK Project
- The Global Change Master Directory
- NASA's EOSDIS Version 0 Information Management System
- NOAA's National Geophysical Data Center/World Data Center—A Marine Geology Inventory Online Search System
- U.S. Geological Survey Node of the National Geospatial Data Clearinghouse, a component of the National Spatial Data Infrastructure (NSDI)

These systems have been chosen as representative of the types of geospatial information retrieval available today. At the end of this discussion there is a list of World Wide Web sites (Appendix A) of interest for geospatial data and information. Again, this is not an exhaustive list but offers several starting points to an exploration of the Web.

The GC-ASK Project

The Global Change Assisted Search for Knowledge (GC-ASK) is a one-year technology development project of the GCDMWG's Access Subgroup. The design, built on commercial off-the-shelf (COTS) products, incorporates the semantic network retrieval engine of ConQuest[6] into a client/server architecture to access distributed databases of all kinds. Plans include the integration of GIS software from E-systems[7] called OASIS (Open Architecture for Scientific Information Systems). There are four prototypes planned in the one-year project. The first prototype demonstrated the ability of the system to search distributed databases that had been indexed by ConQuest software. The second prototype includes a Graphical User Interface, Z39.50 compliance, and additional terminology collections (i.e., National Institutes of Health's Unified Medical Language System [UMLS], Defense Technical Information Center [DTIC] thesaurus, and the NASA Thesaurus). Its debut will be on Earth Day, April 22nd, on the Mall in Washington, D.C.

The third prototype is scheduled to include GIS capability, additional interfaces for multiple user classes, a consistent data presentation model, and connections to non-native (non-ConQuest) search engines to demonstrate the capability to link to and effectively use the functionality of existing search engines of participating agencies. The fourth and final prototype is scheduled to provide the ability to perform assisted searches simultaneously across multiple databases—including non-ConQuest

search engines—and includes retrospective searching, real-time profiling, and on-disk (CD-ROM) product searching. The project will be completed in October 1995.

The GIS component is being contributed to the GC-ASK project by E-systems. They will use their Open Architecture for Scientific Information Systems software product which builds on common industry standards and commercially available software. Core OASIS functions include georeferenced data management using a wide variety of data sources and formats and support for data dissemination, handling, and storage and exploitation for research and analysis. These functions stem from the integration of a geographic information system with image processing and a relational database management system (RDBMS) with a mechanism for incrementally adding data visualization tools. Depending on the capabilities of the host workstation, the OASIS system can display data in a variety of formats and combinations.

The GIS software used in OASIS is Genemap, an open standards-based vector GIS. The functionality envisioned for GC-ASK is a map-based interface that permits zooming and selection of a polygon or point representing the area of interest. The results are displayed graphically, superimposed over the query area as well as in a tabular listing of metadata for the retrieved data sets or documents. Exact implementation of GIS capability in GC-ASK has not been determined at the time that this presentation was written.

GC-ASK is a technology development project—a proof of concept. If it is successful and it is adopted by the GCDIS agencies, users will be able to access both distributed data and information sources with spatial and natural language queries. Spatial access depends on the database objects having geocoding, of course, which is not currently available for the textual records and documents. Another handicap for spatial access to text in the GC-ASK prototype is that the current terminology bases used by ConQuest do not include a geographical thesaurus of place names and their relationships.

The Global Change Master Directory (NASA)

NASA's Global Change Master Directory <http://gcmd.gsfc.nasa.gov> is an international multidisciplinary directory of metadata records for scientific data sets, including satellite and ground observational data, chronological climatological data, and data from global and regional field experiments. It is a major component of GCDIS.

Data sets on individual scientific workstations as well as those in agency data centers are included in the GCMD. The GCMD data descriptions include NASA, NOAA, National Center for Atmospheric Research, USGS, DOE's Carbon Dioxide Information Analysis Center, and EPA data sets, along with entries from universities and research centers. The GCMD

also contains descriptions of data held outside the United States through the International Directory Network (IDN). The GCMD is duplicated in three non-U.S. sites (Italy, Japan, and Canada) and updated every two weeks.

The GCMD database contains approximately 3,000 high-level data set descriptions that give the user basic information on the data sets and points of contact. Through a LINK command, an automatic connection to the site of the data holding can be created, linking directly to external systems for more information, browsing, and even data ordering. These sites currently include the NASA Space Shuttle Earth Observation Program (SSEOP), NOAA National Climatic Data Center (NCDC), USGS Earth Resources Observation Systems Data Center, Global Land Information System (GLIS), CNES SPOT Image Catalog, Earth Observation Satellite Data Inventory Service (SINFONIA), and the University of Rhode Island AVHRR Archive. Future links will include the Earth Observing System Data and Information System and the Consortium for International Earth Science Information Network. The remaining part of the GCMD database encompasses supplemental information on data centers/systems, campaign/projects, spacecraft, and sensors.

GCMD data sets are described in the Data Interchange Format (DIF), which is being modified to comply with the Federal Geographic Data Committee metadata standard. The FGDC metadata standard and the DIF include the following locational data elements:

West_Bounding_Coordinate
East_Bounding_Coordinate
North_Bounding_Coordinate
South_Bounding_Coordinate

These four bounding coordinates can be used to retrieve data sets from the GCMD when the "Query for DIFs" option is chosen from the WWW home page. The other search criteria, such as "Location Keyword," are supported by the lists of controlled vocabulary permitted in the field—the "Valids."

In the WWW interface, there are boxes provided for entering the "Southwestern Coordinate (lat, lon)" and the "Northeastern Coordinate (lat, lon)." The GCMD WWW interface includes a choice of including or excluding the global data sets from the retrieval. The matching algorithms will select any area that overlaps, encloses, or is enclosed by the target area. GCMD can also be accessed through a character interface and an X-Windows interface.

The system response to a query is in the form of a listing of metadata records. The user must take additional steps to obtain access to the data, and contact points are included for this purpose. The GCMD is a direc-

tory level database, meaning that one metadata record can represent a set of data. The data set as a whole may have global coverage, and that is the only geographic level that is provided at the GCMD level. The more detailed level of data coverage—the inventory level—is available from the data center archives.

The WWW interface to the GCMD also includes the option to search the database with the WAIS software. Future plans include improving geographical searches, implementing spatial WAIS, and migrating the system to a "distributed" architecture.

EOSDIS Version 0 Information Management System (NASA)

The Earth Observing System (EOS), part of NASA's Mission to Planet Earth, is NASA's major contribution to the Global Change Research Program. The Data Information System component, EOSDIS, is being designed as a distributed system to support archiving and distribution of data at multiple data centers. The Distributed Active Archive Centers (DAACs) are connected by an Information Management System (IMS) which provides an interface for "one stop shopping" for earth science data, allowing users to search for and order data from multiple data centers in a single session.

Version 0 of EOSDIS, released in September 1994, is a prototype system <http://harp.gsfc.nasa.gov:1729/eosdis_documents/eosdis_home.html>. The interface is available through both graphical and character user interfaces to accommodate a variety of user environments ranging from simple VT100 terminals to sophisticated graphical workstations. It allows the user access to information and metadata through three levels or types of information—directory, guide, and inventory searches. The directory provides concise high-level information about data sets from any point in the system. The guide provides detailed descriptions about data sets, platforms, sensors, projects, and data centers. The inventory search function provides descriptions of specific observations or collections of observations of data (granules) that are available for request from a data archive. Version 0 can be used through any of the NASA DAACs and all of the connections to the DAACs are present on the Version 0 home page. Other functions provided include:

- A Product Request allows users to view information pertaining to orderable data products and then construct a request which is forwarded to the relevant archive for order processing.
- The Coverage Map, available in the graphical version, is a two-dimensional graphical representation of the geographic coverage of selected inventory granules. It displays the Earth in an orthographic projection.

- The Browse function allows a user to locate and retrieve reduced resolution images as an aid to data selection. The user may either view the image in the graphical version IMS interface or have it staged for FTP pick-up.

The Version 0 home page provides links to other home pages that describe various EOSDIS DAACs and cooperative data centers. User support contact information is provided. Also available is a "glossary of terms," an "acronym list," and a "list of known problems."

The geographic search is supported in multiple ways through the X-Windows interface: by selecting from a map, describing a rectangle, by describing four corners of a polygon that do not have right angles and are not parallel to the equator, by describing a point and the height and width of the area around the point, by describing just a point reference, or by global searches. Search results can be limited to only those that are global but it is not possible to eliminate global data sets from the results. This means that a search for any small area will retrieve all global data sets.

Data sets can be selected from the resulting lists of "hits" from each data center based on their descriptions, based on their coverage areas which can be displayed on map projections, or on the basis of browse images when they are available. Orders for the data sets can be entered online.

The EOSDIS IMS software continues to be developed and will be available in an updated version soon. This client software will be available for distribution for those that want to use it locally.

The National Geophysical Data Center (NOAA)

The National Geophysical Data Center (NGDC) <http://www.ngdc.noaa.gov/> manages environmental data in the fields of solar-terrestrial physics, solid earth geophysics, marine geology and geophysics, paleoclimatology, and glaciology (snow and ice). In each of these fields it also operates a World Data Center (WDC) discipline center (from the NGDC home page). The home page points to several sites of interest to a GIS audience. For an illustration of "GIS Techniques Using NOAA Data," choose "Solid Earth Geophysics" and then "GIS Applications using NOAA Data" <http://www.ngdc.noaa.gov/seg/globsys/gisdes.html>. For an example of spatial searching, choose "Marine Geology and Geophysics" <http://www.ngdc.noaa.gov/mgg/mggd.html> and then "Online Searches" and then "SEARCH GEOLIN" <http://www.ngdc.noaa.gov/mgg/geolin/gsearch.HTML>.

"The GEOLIN system contains an inventory of all marine geological data sets and reports available from NGDC in digital or analog form and

supports spatial searching via the 'latest' forms interface of X-Mosaic or Mosaic for Windows" (from the home page). The GEOLIN search uses Mosaic forms to launch an experimental version of cross-platform libraries developed at NGDC in cooperation with DataWare Technologies, Inc. of Boston, Massachusetts, and Boulder, Colorado, based on the Record ReferenceBook libraries owned by DataWare Technologies, Inc. (from the GEOLIN introductory page).

In this forms-based spatial searching interface, users are prompted for upper and lower latitude and left and right longitude. The minus sign is used for southern latitudes and western longitudes, a common device. All overlapping areas are retrieved no matter how broad the coverage.

Also from the National Geophysical Data Center in Boulder, Colorado <http://www.ngdc.noaa.gov/seg/globsys/global_c.html> are global environmental data sets on CD-ROM produced by the Global Ecosystems Database Project, conducted by NGDC and the U.S. Environmental Protection Agency's Environmental Research Laboratory in Corvallis, Oregon. The databases contain raster gridded map layers registered to a common latitude-longitude base. Each file has been inspected for optimal quality and usability for analysis based on the current state of data development. Parameters have been chosen for their potential use in integrated studies of the global environment. Individual data layers have been contributed by many scientific laboratories and individuals. More information about these CD-ROMs is included in Appendix C of this paper.

The NOAA Data Set Catalog <http://www.esdim.noaa.gov/NOAA-Catalog/NOAA-Catalog.html> provides word searching, phrase and root searching, Boolean and field-based searching of its metadata records, and gives a generous set of information to help users format and conduct their searches. They do not have a spatial component currently but include "latitude/longitude searches" in their list of future plans.

National Geospatial Data Clearinghouse (USGS)

The USGS node of the National Geospatial Data Clearinghouse is a component of the National Spatial Data Infrastructure <http://nsdi.usgs.gov/nsdi/index.html>. Sets of USGS geospatial data can be searched by keyword or spatial extent using SpatialWAIS. Data sets are described using the Federal Geographic Data Committee Content Standards for Digital Geospatial Metadata.

Browsing and searching capabilities are available. By using the search option, you can identify your interest by keyword (a WAIS search) or you can identify an area by its latitude and longitude boundaries using the forms interface to WAIS. Another option is to choose a state or states as the area.

Once you find a data set you want, its metadata files contain instructions for obtaining the data set. Some smaller data sets are available online at no cost. For others, you may need to set up an account to pay a small distribution fee. A few data sets are not online, and you will be instructed how to order a compact disk or a tape.

Many USGS data products are available in the Spatial Data Transfer Standard (SDTS) format, a new federal standard to ensure data compatibility, or in the older Digital Line Graph (DLG) format. Others are in more specialized formats, which are described in their metadata. Whenever you retrieve a data set, it is important to retrieve the metadata as well, since the metadata provides important information for using the data set.

The ranking algorithm that is operating in this system is based only on the words entered for the WAIS search—the geographic parameters are not incorporated. Therefore, if you enter a search composed only of spatial parameters, the ranking that is returned is purely arbitrary.

The USGS is moving to the I-Site software from the Clearinghouse for Networked Information Discovery and Retrieval (CNIDR) (Nebert, 1995). I-Site is based on version 2 of the Z39.50 protocol with the incorporation of some elements from version 3; it has field-based search capability as well as the means by which to call other search software. A description of I-Site can be located at <http://vinca.cnidr.org/software/Isite/Isite.html>.

SUMMING UP—WHAT IS THE STATUS OF SPATIAL DISCOVERY AND DISPLAY TOOLS FOR GLOBAL CHANGE DATA?

The spatial access tools for the Internet environment are in the developmental stages. They show exciting promise for the near future as additional tools for network discovery of data relevant to the study of a particular geographic area. More visual map-based interfaces are appearing that support the representation of areas. Most systems, however, are still limited to the representation of regular polygons or points. Matching algorithms to return a ranked list on the basis of geographic similarity are not being used—at least not in the systems that I reviewed.

Spatial access to textual information along with data sets is an area that needs more emphasis through research and development to build translations between the geographic text and geocoded representations. There are several efforts in this direction. The GeoWeb Project is being developed by Brandon Plewe of the State University of New York at Buffalo (plewe@acsu.buffalo.edu) and Chris Stuber of the U.S. Census Bureau (Chris.Stuber@census.gov) (Plewe, 1994). The concept of GeoWeb

includes a gazetteer interface to expand geographic terminology to a latitude/longitude point reference that can be used for searching. The second is the GIPSY project by Allison Woodruff and Christian Plaunt (1994) at the University of California, Berkeley, which processes the natural language of the document with a "spatial reasoning" algorithm "to approximate statistically the geoposition being referenced in the text." GIPSY uses not only place name terms from the text but also terms that describe geologic features, climate, land use, habitats, and area size and reference terminology sets such as the USGS Geographic Names Information System and the USGS Geographic Information Retrieval and Analysis System (GIRAS). In addition, the NSF/ARPA/NASA digital library project at the University of California, Berkeley, is planning to develop a gazetteer-spatial link for their project.

There is potential for users seeking georeferenced information to discover and obtain the information they need. Fortunately, many free sources of information exist—many open access systems are being built. However, most users are going to need help—the tools are not standardized enough, the systems are not interoperable enough, the coverage is not complete enough, and there is too much to keep up with in this rapidly changing environment. This help should be available through library services that specialize in geographic information as well as specific governmental offices.

GCDIS, in particular, provides an avenue for librarians to participate in the building and expansion of the systems to access global change data and information, including spatial data and information. Through the Library Information Subgroup, librarians can directly participate in evaluation projects and voice their needs and ideas in an effective way. In particular, librarians are invited to participate in the evaluation of the GCASK retrieval software engineering project, in the evaluation of the GCDIS gopher and home page, and in the evaluation of the systems available through the participating federal agencies.

APPENDIX A. INTERNET SITES FOR THE SYSTEMS DISCUSSED AND RELATED SITES

Board of Geographic Names
http://toponym.dma.gov/check_login.html

CIESIN
http://www.ciesin.org

EOSDIS
http://harp.gsfc.nasa.gov:1729/eosdis_documents/eosdis_home.html

Fish and Wildlife Service Hotlist
http://www.fws.gov/hotlist.html#Geospatial

GC-ASK (Global Change - Assisted Search for Knowledge Project)
telnet://esdim2.edim.noaa.gov
Login id is "conquest"
Password is "conquest11"

For more information about the project, send e-mail to gcdis.lis@earth.usgcrp.gov.

GCDIS (Global Change Data and Information System)
gopher://gopher.gcdis.usgcrp.gov
If login required: "gopher"
http://www.gcdis.usgcrp.gov (under construction)
E-mail: gcdis.lis@earth.usgcrp.gov

GCMD (Global Change Master Directory)
http://gcmd.gsfc.nasa.gov

GCRIO (Global Change Research Information Office)
http://www.gcrio.org
gopher://gopher.gcrio.org
E-mail: help@gcrio.org

GEONET-L (Geoscience Librarians & Information Specialists)
geonet-l@iubvm.ucs.indiana.edu

GeoWeb Project (description included in Appendix C)
http://wings.buffalo.edu/geoweb/general.html

Ionia "1 km AVHRR Global Land Data Set" Net-Browser (European Space Agency, Italy)
http://shark1.esrin.esa.it/

LASR Project (Library Access, Search, and Retrieval Project to evaluate GCDIS)
http://juliet.cs.virginia.edu/lasr/

Marc's Favorite Internet Sites (Marc E. Berryman, marc.berryman@dir.state.tx.us)
http://www.texas.gov/DIR/people/mab.html

National Geospatial Data Clearinghouse. U.S. Geological Survey Node
http://nsdi.usgs.gov/nsdi/index.html

NOAA Marine Geology & Geophysics On-line Searches
http://www.ngcd.noaa.gov/mgg/geolsys.html

Sol Katz List of Spatial Geographic Search Sites
ftp://ftp.blm.gov/pub/gis/nsdi.html

TMS Experimental Browser (description included in Appendix C)
http://tiger.census.gov/

APPENDIX B. ACRONYMS

ARM	Atmospheric Radiation Measurement
ARS	Agricultural Research Service, USDA
AVHRR	Advanced Very High Resolution Radiometer
CDIAC	Carbon Dioxide Information Analysis Center, DOE
CIESIN	Consortium for International Earth Science Information Network, NASA
CSREES	Cooperative State Research, Education and Extension Service, USDA
CDDIS	Crustal Dynamics Data Information System, NASA
DAAC	Distributed Active Archive Center
DIF	Directory Interchange Format, GCMD
DOE	Department of Energy
DOI	Department of Interior
DTIC	Defense Technical Information Center, DOD
EIA	Energy Information Administration, DOE
EOSDIS	Earth Observation System Data and Information System, NASA
EPA	Environmental Protection Agency
EREN	Energy Efficiency and Renewable Energy Network, DOE
EROS	Earth Resources Observation Systems, USGS
ERS	Economic Research Service, USDA
ESA	European Space Agency
ESDIS	Earth Science Data and Information System
FGDC	Federal Geographic Data Committee
FS	Forest Service, USDA
GC-ASK	Global Change Assisted Search for Knowledge Project
GCDIS	Global Change Data and Information System
GCDMWG	Global Change Data Management Working Group
GCMD	Global Change Master Directory
GCRIO	Global Change Research Information Office
GCRP	Global Change Research Program
GILS	Government Information Locator Service
GIS	Geographic Information System
GUI	Graphical User Interface
IWGDMGC	Interagency Working Group on Data Management for Global Change (previous name for GCDMWG)
LASR	Library, Access, Search, and Retrieval Project, LIS
LIS	Library Information Subgroup, GCDMWG
NAL	National Agricultural Library, USDA
NASA	National Aeronautics and Space Administration
NBS	National Biological Service, DOI
NCAR	National Center for Atmospheric Research, NSF
NEDI	National Environmental Data Inventory
NGDC	National Geophysical Data Center
NOAA	National Oceanic and Atmospheric Administration, Department of Commerce
NODIS	NASA Online Data and Information System
NSF	National Science Foundation
NSDI	National Spatial Data Infrastructure

NSSDC	NASA Space Science Data Center
NTIS	National Technical Information Service (NTIS)
OASIS	Open Architecture for Scientific Information Systems, E-Systems
ORNL	Oak Ridge National Laboratory, DOE
OSTI	Office of Scientific and Technical Information, DOE
TERRA	Terrestrial Ecosystems Regional Research and Analysis Laboratory, USDA
UMLS	Unified Medical Language System
USDA	U.S. Department of Agriculture
USGS	U.S. Geological Survey
WAOB	World Agricultural Outlook Board, USDA
WWW	World Wide Web

APPENDIX C. NATIONAL GEOPHYSICAL DATA CENTER, BOULDER CO. GLOBAL CHANGE CD-ROMS (ADAPTED FROM HOME PAGE)

Also from the National Geophysical Data Center in Boulder, Colorado <http://www.ngdc.noaa.gov/seg/globsys/global_c.html>, are global environmental data sets on CD-ROM produced by the Global Ecosystems Database Project, conducted by NGDC and the U.S. Environmental Protection Agency's Environmental Research Laboratory, in Corvallis, Oregon. The databases contain raster gridded map layers registered to a common latitude-longitude base. Each file has been inspected for optimal quality and usability for analysis given the current state of data development. Parameters have been chosen for their potential use in integrated studies of the global environment. Individual data layers have been contributed by many scientific laboratories and individuals.

"Disc A" includes selected data on the global environment, such as ecosystems, land use, wetlands, vegetation (including satellite-derived vegetation index), climate, topography, and soils. These data are on a range of compatible grids, from 2 degrees to 2 minutes. Vector data for coastlines and other features are also provided.

"Disc A," in beta test since last year, is readable on IBM DOS-compatible machines, Apple Macintoshes, UNIX workstations, and other computers that support the ISO 9660 standard. Features of the Global Ecosystems Data on CD-ROM:

- *Integration of Data for Multivariate Analysis.* The data have undergone extensive quality control and preparation of documentation (300+ pages) for improved data synergy and usefulness for integrated multivariate spatial analysis. Several new techniques have been developed in this process. This process included format and grid conversion, registration and/or reprojection, creating documentation as needed, etc., while preserving the original nature and content of the data.
- *Scientific Peer Review of Data.* From NGDC's network of colleagues, cooperative projects with the International Geosphere-Biosphere Programme and the U.S. Environmental Protection Agency, over 100 scientists have reviewed the data. This review, based on prototypes since 1989 and a full beta test of a prototype CD-ROM in 1991-1992, has resulted in significant improvements to design, documentation, and content. Peer review comments are summarized with the database. This process continues, supporting future updates.
- *Public Domain, Accessible.* The data are completely within the public domain, per agreement by authors of each data set. All "value-added" work performed at NGDC in integrating the data is also in the public domain.
- *Data and Software Support Research, Education, Awareness—Optimized for use on Geographic Information Systems.* We provide software to browse, visualize, and select appropriate portions of the database for your work. Primarily, however, the database is designed to interface with many statistical, image analysis, and geographic information systems. Data and sample setups are provided in two highly accessible geographic information systems—GRASS and IDRISI—to show how they can be used in other GISs. The structure of the database has demonstrated its adaptability by passing several tests by software vendors for quick and easy importation into their systems.

- *Provide Feedback to Authors of Data Sets.* The process of peer review and integration includes continued interaction with authors of individual data sets. Data enhancements and corrections are made cooperatively with creators of the data sets. The feedback that we obtain in our normal duties as a Data Center also helps to provide ideas and resources for improving data sets.

There is also a beta version of "Disc B" which contains:

Global Vegetation Index data produced at Chiba University, Japan
Weekly sea surface temperature
Derived snow cover and snow depth data
Ecoregions data
Improved urban coverage data
Elevation and bathymetry
Forest practices database
EPA climate database and doubled CO2 predictions for the US
Potential natural vegetation of the conterminous US
US boundaries from Micro World Data Bank II
EPA rice climatology—UV-B irradiance and agroclimatology for South, Southeast, and East Asia

Other global and regional data sets available from NGDC:

- *Experimental Calibrated Global Vegetation Index on CD-ROM.* Data from NOAA's Advanced Very High Resolution Radiometer. Originally designed and produced by Kevin Gallo of the Office of Research and Applications of NOAA's National Environmental Satellite, Data and Information Service. Please request flier SE-2008 for more information about this data set.
- *Global Change Education Diskette Project.* Originally developed for the International Geosphere-Biosphere Programme, this is an integrated Global Change Data Base for Africa, with accompanying documentation and a workbook of exercises designed for classroom and self-instruction in using Geographic Information Systems to explore the global or continental environment. Please request flier SE-2007 for more information about this data set.

How to Order:

The price for the Global Ecosystems Data on CD-ROM, with browse and visualization software and extensive documentation is $101 (product number 1016-A27-001). Data contributors and academic researchers should contact NGDC for information about obtaining data by special arrangement.

Make checks and money orders payable to COMMERCE/NOAA/NGDC. Do not send cash. All non-U.S.A. orders must be in U.S. Dollars drawn on a U.S.A. bank. A ten-dollar ($10) handling fee is required for delivery outside the U.S.A. Orders may be charged to American Express, MasterCard, or VISA by telephone, letter, or fax. Please include a credit card account number, expiration date, telephone number, and your signature with order. Address inquiries and orders to:

National Geophysical Data Center
NOAA, E/GC1, Dept 891
325 Broadway
Boulder, CO 80303, U.S.A.
Phone: 303-497-6900

Fax: 303-497-6513
Internet: info@mail.ngdc.noaa.gov
Telex: 592811 NOAA MASC BDR

For questions about Global Ecosystems Data on CD-ROM please contact: John Kineman, 303-497-6900, jjk@mail.ngdc.noaa.gov

THE GEOWEB PROJECT

The GeoWeb project <http://wings.buffalo.edu/geoweb/general.html> was started in May 1994 and is in the developmental stage. It is being spearheaded by Brandon Plewe of the State University of New York at Buffalo as his project for an M.A. in Geography (plewe@acsu.buffalo.edu) and Chris Stuber of the U.S. Census Bureau (Chris.Stuber@census.gov) as part of his work in building better interfaces to census data. The mission of GeoWeb is to provide a service which geographers, GIS users, and the general public can use to locate geographic information that has been made available on the Internet. To fulfill this mission, GeoWeb consists of two parts: an index of Internet-accessible geographic information and one or more WWW interfaces to this index.

Internet resources in the index will include those for the general public (e.g., general information about countries, states, and places; simple maps of areas; lists and maps of Internet resources in an area) and those for cartographers and geographers (e.g., cartographic/GIS base map files, thematic data of a geographic nature, and GIS data sets).

This index will contain metadata (based on FGDC standards), giving a brief description of each data set, and a URL of its location on the net. The plan is to have the index be at the data set (not the series) level; for instance, each 1degx1deg DEM would have a separate entry (several hundred for the entire series). A separate index of more complete metadata for an entire series or URLs to series metadata pages on the WWW could be included in the future. The first phase will only include data for the United States, but the goal is to make it scalable for worldwide metadata.

The database will probably be based on the emerging Spatial WAIS search engine. However, it will need useful gateways to make it accessible to the public. Four possibilities have been discussed:

1. *Direct WAIS search.* A form with several fields would be presented (including spatial), which the user could fill out and submit. A page would be returned listing the resources which matched the criteria with hypertext links for each entry to more complete metadata and the resource itself.
2. *Dynamic map window.* A WWW page displays a simple map which can be scrolled and zoomed (a la the Xerox Map Server). At any time, the user could select a link to see a list of the resources that cover the displayed region.
3. *Gazetteer.* Based on a U.S. Gazetteer. The user would do a name search for a place (state, county, city, etc.), receiving in return a brief description of the place (from GNIS or census data) and a list of the resources that cover that place.

4. *Hierarchical Directory.* Based on either a menu or map format (similar to the Virtual Tourist). The user would first get a directory of states (or countries). Selecting one would bring up a new page containing links to a list of statewide resources and the counties in the state. Selecting a county would display a page of county-wide resources and places in the county. Selecting a place would display a list of resources covering that place.

A paper by Brandon Plewe (1994) on the GeoWeb Project can be accessed at <http://wings.buffalo.edu/~plewe/paperwww.html>

TMS EXPERIMENTAL BROWSER

Brandon Plewe has also created a program called the TigerWWW for the Census Bureau's TMS Experimental Browser <http://tiger.census.gov/>. This site is the Washington, DC area from TIGER/92 data. Other preset values are The Mall in Washington, DC, the United States, Northeast U.S., and New York City. The interface prompts for a point location (a latitude and longitude) and a width and height in degrees for the map. Zooming in and out, specifying a point, and placing markers are all supported for the displayed map. This is an example of access to a particular set of cartographic data by the spatial description of a point and an area around that point. The approach shows promise.

NOTES

[1] This group is also known as the Interagency Working Group on Data Management for Global Change (IWGDMGC).

[2] For more information about the activities of the Library Information Subgroup, send email to gcdis.lis@earth.usgcrp.gov.

[3] GRASS (Geographical Resources Support System) is public domain software developed by the U.S. Army Construction Engineering Research Laboratory (USA-CERL) in Champaign, Illinois.

[4] GeoRef is a database created by the American Geological Institute that covers the worldwide literature on geology and geosciences. It is available through various database vendors and on CD-ROM.

[5] An interesting project along this line is the NASA-funded project entitled "Compression and Progressive Transmission of Digital Images." The principal investigator is Dr. Jeffrey W. Percival of the University of Wisconsin-Madison who can be reached at jwp@bernie.sal.wisc.edu.

[6] Semantic net technology uses dictionary definitions and thesaurus links to build relationships—both word-form and meaning-based relationships—between words so that the words of a query can be expanded into related terminology for the search of the database. ConQuest Software is located at 9700 Patuxent Woods Drive, Suite 140, Columbia, MD 21046.

[7] The address of E-Systems Inc. is Garland Division, PO Box 660023, Dallas, TX 75266-0023.

REFERENCES

Ben-Avraham, A.; Kempler, D.; & Ginzburg, A. (1988). Plate convergence in the Cyprean Arc. *Tectonophysics, 146,* 231-240.

Committee on the Environment and Natural Resources (draft of July 7, 1994). *The U.S. Global Change Data and Information System Implementation Plan.*

Hill, L. L. (1990). *Access to geographic concepts in online bibliographic files: Effectiveness of current practices and the potential of a graphic interface.* Unpublished dissertation, School of Library and Information Science, University of Pittsburgh, December.

Nebert, D. D. (1995). *Use of Z39.50 to search and retrieve geospatial data.* Manuscript submitted for publication. Preprint available from http://130.11.51.187/public/DLIpaper395.html

Plewe, B. (1994). *The GeoWeb project: Using WAIS and the World Wide Web to aid location of distributed data sets.* Available from http://wings.buffalo.edu/~plewe/paperwww.html

Woodruff, A. G., & Plaunt, C. (1994). GIPSY: Automated geographic indexing to text documents. *Journal of the American Society for Information Science,* 45(9), 645-655.